

Pure Strategy or Mixed Strategy?

Jun He*

Department of Computer Science
Aberystwyth University, Aberystwyth, SY23 3DB, U.K.
jun.he@ieee.org

Feidun He

School of Information Science and Technology, Southwest Jiaotong University
Chengdu, Sichuan, 610031, China

Hongbin Dong

College of Computer Science and Technology, Harbin Engineering University
Harbin, 150001, China

January 30, 2012

Abstract

Mixed strategy evolutionary algorithms (EAs) aim at integrating several mutation operators into a single algorithm. However no analysis has been made to answer the theoretical question: whether and when is the performance of mixed strategy EAs better than that of pure strategy EAs? In this paper, asymptotic convergence rate and asymptotic hitting time are proposed to measure the performance of EAs. It is proven that the asymptotic convergence rate and asymptotic hitting time of any mixed strategy (1+1) EA consisting of several mutation operators is not worse than that of the worst pure strategy (1+1) EA using only one mutation operator. Furthermore it is proven that if these mutation operators are mutually complementary, then it is possible to design a mixed strategy (1+1) EA whose performance is better than that of any pure strategy (1+1) EA using only one mutation operator.

Keywords: Mixed Strategy, Pure Strategy, Asymptotic Convergence Rate, Asymptotic Hitting Time, Hybrid Evolutionary Algorithms

1 Introduction

Different search operators have been proposed and applied in EAs [1]. Each search operator has its own advantage. Therefore an interesting research issue is to combine the advantages of variant operators together and then design more efficient hybrid EAs. Currently hybridization of evolutionary algorithms becomes popular due to their capabilities in handling some real world problems [2].

Mixed strategy EAs, inspired from strategies and games [3], aims at integrating several mutation operators into a single algorithm [4]. At each generation, an individual will choose one mutation operator according to a strategy probability distribution. Mixed strategy evolutionary programming has been implemented for continuous optimization and experimental results show it performs better than its rival, i.e., pure strategy evolutionary programming which utilizes a single mutation operator [5, 6].

However no analysis has been made to answer the theoretical question: whether and when is the performance of mixed strategy EAs better than that of pure strategy EAs? This paper aims at providing an initial answer. In theory, many of EAs can be regarded as a matrix iteration procedure. Following matrix iteration analysis [7], the performance of EAs is measured by the asymptotic convergence rate, i.e., the spectral radius of a probability transition sub-matrix associated with an EA. Alternatively the performance of EAs can be measured by the asymptotic hitting time [8], which approximatively equals the reciprocal of the asymptotic

*Corresponding author.

convergence rate. Then a theoretical analysis is made to compare the performance of mixed strategy and pure strategy EAs.

The rest of this paper is organized as follows. Section 2 describes pure strategy and mixed strategy EAs. Section 3 defines asymptotic convergence rate and asymptotic hitting time. Section 4 makes a comparison of pure strategy and mixed strategy EAs. Section 5 concludes the paper.

2 Pure Strategy and Mixed Strategy EAs

Before starting a theoretical analysis of mixed strategy EAs, we first demonstrate the result of a computational experiment.

Example 1. *Let's see an instance of the average capacity 0-1 knapsack problem [9, 10]:*

$$\begin{aligned} & \text{maximize } \sum_{i=1}^{10} v_i b_i, & b_i \in \{0, 1\}, \\ & \text{subject to } \sum_{i=1}^{10} w_i b_i \leq C, \end{aligned} \quad (1)$$

where $v_1 = 10$ and $v_i = 1$ for $i = 2, \dots, 10$; $w_1 = 9$ and $w_i = 1$ for $i = 2, \dots, 10$; $C = 9$.

The fitness function is that for $x = (b_1, \dots, b_{10})$

$$f(x) = \begin{cases} \sum_{i=1}^{10} v_i b_i, & \text{if } \sum_{i=1}^{10} w_i b_i \leq C, \\ 0, & \text{if } \sum_{i=1}^{10} w_i b_i > C. \end{cases}$$

We consider two types of mutation operators:

- *s1: flip each bit b_i with a probability 0.1;*
- *s2: flip each bit b_i with a probability 0.9;*

The selection operator is to accept a better offspring only.

Three (1+1) EAs are compared in the computation experiment: (1) EA(s1) which adopts s1 only, (2) EA(s2) with s2 only, and (3) EA(s1,s2) which chooses either s1 or s2 with a probability 0.5 at each generation.

Each of these three EAs runs 100 times independently. The computational experiment shows that EA(s1, s2) always finds the optimal solution more quickly than other twos.

This is a simple case study that shows a mixed strategy EA performs better than a pure strategy EA. In general, we need to answer the following theoretical question: whether or when do a mixed strategy EAs are better than pure strategy EAs?

Consider an instance of the discrete optimization problem which is to maximize an objective function $f(x)$:

$$\max\{f(x); x \in S\}, \quad (2)$$

where S a finite set. For the analysis convenience, suppose that all constraints have been removed through an appropriate penalty function method. Under this scenario, all points in S are viewed as feasible solutions. In evolutionary computation, $f(x)$ is called a *fitness function*.

The following notation is used in the algorithm and text thereafter.

- $x, y, z \in S$ are called *points* in S , or *individuals* in EAs or *states* in Markov chains.
- The *optimal set* $S_{\text{opt}} \subseteq S$ is the set consisting of all optimal solutions to Problem (2) and *non-optimal set* $S_{\text{non}} := S \setminus S_{\text{opt}}$.
- t is the generation counter. A random variable Φ_t represents the state of the t -th generation parent; $\Phi_{t+1/2}$ the state of the child which is generated through mutation.

The mutation and selection operators are defined as follows:

- A *mutation operator* is a probability transition from S to S . It is defined by a *mutation probability transition matrix* \mathbf{P}_m whose entries are given by

$$P_m(x, y), \quad x, y \in S. \quad (3)$$

- A *strict elitist selection operator* is a mapping from $S \times S$ to S , that is for $x \in S$ and $y \in S$,

$$z = \begin{cases} x, & \text{if } f(y) \leq f(x), \\ y, & \text{if } f(y) > f(x). \end{cases} \quad (4)$$

A *pure strategy* (1+1) EA, which utilizes only one mutation operator, is described in Algorithm 1.

Algorithm 1 Pure Strategy Evolutionary Algorithm EA(s)

```

1: input: fitness function;
2: generation counter  $t \leftarrow 0$ ;
3: initialize  $\Phi_0$ ;
4: while stopping criterion is not satisfied do
5:    $\Phi_{t+1/2} \leftarrow$  mutate  $\Phi_t$  by mutation operator s;
6:   evaluate the fitness of  $\Phi_{t+1/2}$ ;
7:    $\Phi_{t+1} \leftarrow$  select one individual from  $\{\Phi_t, \Phi_{t+1/2}\}$  by strict elitist selection;
8:    $t \leftarrow t + 1$ ;
9: end while
10: output: the maximal value of the fitness function.
```

The stopping criterion is that the running stops once an optimal solution is found. If an EA cannot find an optimal solution, then it will not stop and the running time is infinite. This is common in the theoretical analysis of EAs.

Let s_1, \dots, s_κ be κ mutation operators (called *strategies*). Algorithm 2 describes the procedure of a *mixed strategy* (1+1) EA. At the t -th generation, one mutation operator is chosen from the κ strategies according to a *strategy probability distribution*

$$q_{s_1}(x), \dots, q_{s_\kappa}(x), \quad (5)$$

subject to $0 \leq q_s(x) \leq 1$ and $\sum_s q_s(x) = 1$.

Write this probability distribution in short by a vector $\mathbf{q}(x) = [q_s(x)]$.

Algorithm 2 Mixed Strategy Evolutionary Algorithm EA(s_1, \dots, s_κ)

```

1: input: fitness function;
2: generation counter  $t \leftarrow 0$ ;
3: initialize  $\Phi_0$ ;
4: while stopping criterion is not satisfied do
5:   choose a mutation operator  $s_k$  from  $s_1, \dots, s_\kappa$ ;
6:    $\Phi_{t+1/2} \leftarrow$  mutate  $\Phi_t$  by mutation operator  $s_k$ ;
7:   evaluate  $\Phi_{t+1/2}$ ;
8:    $\Phi_{t+1} \leftarrow$  select one individual from  $\{\Phi_t, \Phi_{t+1/2}\}$  by strict elitist selection;
9:    $t \leftarrow t + 1$ ;
10: end while
11: output: the maximal value of the fitness function.
```

Pure strategy EAs can be regarded a special case of mixed strategy EAs with only one strategy. EAs can be classified into two types:

- A *homogeneous EA* is an EA which applies the same mutation operators and same strategy probability distribution for all generations.
- An *inhomogeneous EA* is an EA which doesn't apply the same mutation operators or same strategy probability distribution for all generations.

This paper will only discuss *homogeneous EAs* mainly due to the following reason:

- The probability transition matrices of an inhomogeneous EA may be chosen to be totally different at different generations. This makes the theoretical analysis of an inhomogeneous EA extremely hard.

3 Asymptotic Convergence Rate and Asymptotic Hitting Time

Suppose that a homogeneous EA is applied to maximize a fitness function $f(x)$, then the population sequence $\{\Phi_t, t = 0, 1, \dots\}$ can be modelled by a *homogeneous Markov chain* [11, 12]. Let \mathbf{P} be the probability transition matrix, whose entries are given by

$$P(x, y) = P(\Phi_{t+1} = y \mid \Phi_t = x), \quad x, y \in S.$$

Starting from an initial state x , the mean number $m(x)$ of generations to find an optimal solution is called the *hitting time* to the set S_{opt} [13].

$$\begin{aligned} \tau(x) &:= \min\{t; \Phi_t \in S_{\text{opt}} \mid \Phi_0 = x\}, \\ m(x) &:= E[\tau(x)] = \sum_{t=0}^{+\infty} tP(\tau(x) = t). \end{aligned}$$

Let's arrange all individuals in the order of their fitness from high to low: x_1, x_2, \dots , then their hitting times are:

$$m(x_1), m(x_2), \dots$$

Denote it in short by a vector $\mathbf{m} = [m(x)]$.

Write the transition matrix \mathbf{P} in the canonical form [14],

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ * & \mathbf{T} \end{pmatrix}, \quad (6)$$

where \mathbf{I} is a unit matrix and $\mathbf{0}$ a zero matrix. \mathbf{T} denotes the probability transition sub-matrix among non-optimal states, whose entries are given by

$$P(x, y), \quad x \in S_{\text{non}}, y \in S_{\text{non}}.$$

The part $*$ plays no role in the analysis.

Since $\forall x \in S_{\text{opt}}, m(x) = 0$, it is sufficient to consider $m(x)$ on non-optimal states $x \in S_{\text{non}}$. For the simplicity of notation, the vector \mathbf{m} will also denote the hitting times for all non-optimal states: $[m(x)], x \in S_{\text{non}}$.

The Markov chain associated with an EA can be viewed as a matrix iterative procedure, where the iterative matrix is the probability transition sub-matrix \mathbf{T} . Let \mathbf{p}_0 be the vector $[p_0(x)]$ which represents the probability distribution of the initial individual:

$$p_0(x) := P(\Phi_0 = x), \quad x \in S_{\text{non}},$$

and \mathbf{p}_t the vector $[p_t(x)]$ which represents the probability distribution of the t -generation individual:

$$p_t(x) := P(\Phi_t = x), \quad x \in S_{\text{non}}.$$

If the spectral radius $\rho(\mathbf{T})$ of the matrix \mathbf{T} satisfies: $\rho(\mathbf{T}) < 1$, then we know [7]

$$\lim_{t \rightarrow \infty} \|\mathbf{p}_t\| = 0.$$

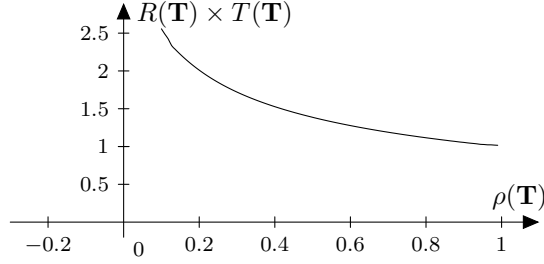
Following matrix iterative analysis [7], the asymptotic convergence rate of an EA is defined as below.

Definition 1. *The asymptotic convergence rate of an EA for maximizing $f(x)$ is*

$$R(\mathbf{T}) := -\ln \rho(\mathbf{T}) \quad (7)$$

where \mathbf{T} is the probability transition sub-matrix restricted to non-optimal states and $\rho(\mathbf{T})$ its spectral radius.

Figure 1: The relationship between the asymptotic hitting time and asymptotic convergence rate: $1/R(\mathbf{T}) < T(\mathbf{T}) < 1.5/R(\mathbf{T})$ if $\rho(\mathbf{T}) \geq 0.5$.



Asymptotic convergence rate is different from previous definitions of convergence rate based on matrix norms or probability distribution [12].

Note: Asymptotic convergence rate depends on both the probability transition sub-matrix \mathbf{T} and fitness function $f(x)$. Because the spectral radius of the probability transition matrix $\rho(\mathbf{P}) = 1$, thus $\rho(\mathbf{P})$ cannot be used to measure the performance of EAs. Because the mutation probability transition matrix is the same for all functions $f(x)$, and $\rho(\mathbf{P}_m) = 1$, so $\rho(\mathbf{P}_m)$ cannot be used to measure the performance of EAs too.

If $\rho(\mathbf{T}) < 1$, then the hitting time vector satisfies (see Theorem 3.2 in [14]),

$$\mathbf{m} = (\mathbf{I} - \mathbf{T})^{-1}\mathbf{1}. \quad (8)$$

The matrix $\mathbf{N} := (\mathbf{I} - \mathbf{T})^{-1}$ is called the *fundamental matrix* of the Markov chain, where \mathbf{T} is the probability transition sub-matrix restricted to non-optimal states.

The spectral radius $\rho(\mathbf{N})$ of the fundamental matrix can be used to measure the performance of EAs too.

Definition 2. The asymptotic hitting time of an EA for maximizing $f(x)$ is

$$T(\mathbf{T}) = \begin{cases} \rho(\mathbf{N}) = \rho((\mathbf{I} - \mathbf{T})^{-1}), & \text{if } \rho(\mathbf{T}) < 1, \\ +\infty, & \text{if } \rho(\mathbf{T}) = 1. \end{cases}$$

where \mathbf{T} is the probability transition sub-matrix restricted to non-optimal states and \mathbf{N} is the fundamental matrix.

From Lemma 5 in [8], we know the asymptotic hitting time is between the best and worst case hitting times, i.e.,

$$\min\{m(x); x \in S_{\text{non}}\} \leq T(\mathbf{T}) \leq \max\{m(x); x \in S_{\text{non}}\}. \quad (9)$$

From Lemma 3 in [8], we know

Lemma 1. For any homogeneous $(1+1)$ -EA using strictly elitist selection, it holds

$$\begin{aligned} \rho(\mathbf{T}) &= \max\{P(x, x); x \in S_{\text{non}}\}, \\ \rho(\mathbf{N}) &= \frac{1}{1 - \rho(\mathbf{T})}, \quad \text{if } \rho(\mathbf{T}) < 1. \end{aligned}$$

From Lemma 1 and Taylor series, we get that

$$R(\mathbf{T})T(\mathbf{T}) = \sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{1}{T(\mathbf{T})} \right)^{k-1}.$$

If we make a mild assumption $T(\mathbf{T}) \geq 2$, (i.e., the asymptotic hitting time is at least two generations), then the asymptotic hitting time approximatively equals the reciprocal of the asymptotic convergence rate (see Figure 1).

Example 2. Consider the problem of maximizing the One-Max function:

$$f(x) = |x|,$$

where $x = (b_1 \dots b_n)$ a binary string, n the string length and $|x| := \sum_{i=1}^n b_i$. The mutation operator used in the $(1+1)$ EA is to choose one bit randomly and then flip it.

Then asymptotic convergence rate and asymptotic hitting time are

$$\begin{aligned} 1/n < R(\mathbf{T}) < 1/(n-1), \\ T(\mathbf{T}) &= n. \end{aligned}$$

4 A Comparison of Pure Strategy and Mixed Strategy

In this section, subscripts \mathbf{q} and \mathbf{s} are added to distinguish between a mixed strategy EA using a strategy probability distribution \mathbf{q} and a pure strategy EA using a pure strategy \mathbf{s} . For example, $\mathbf{T}_{\mathbf{q}}$ denotes the probability transition sub-matrix of a mixed strategy EA; $\mathbf{T}_{\mathbf{s}}$ the transition sub-matrix of a pure strategy EA.

Theorem 1. Let s_1, \dots, s_{κ} be κ mutation operators.

1. The asymptotic convergence rate of any mixed strategy EA consisting of these κ mutation operators is not smaller than the worst pure strategy EA using only one of these mutation operator;
2. and the asymptotic hitting time of any mixed strategy EA is not larger than the worst pure strategy EA using one only of these mutation operator.

Proof. (1) From Lemma 1 we know

$$\begin{aligned} \rho(\mathbf{T}_{\mathbf{q}}) &= \max\left\{\frac{1}{\kappa} \sum_{k=1}^{\kappa} P_{s_k}(x, x); x \in S_{\text{non}}\right\} \\ &\leq \frac{1}{\kappa} \sum_{k=1}^{\kappa} \rho(\mathbf{T}_{s_k}) \\ &\leq \max\{\rho(\mathbf{T}_{s_k}); k = 1, \dots, \kappa\}. \end{aligned}$$

Thus we get that

$$R(\mathbf{T}_{\mathbf{q}}) := -\ln \rho(\mathbf{T}_{\mathbf{q}}) \geq \max\{-\ln \rho(\mathbf{T}_{s_k}); k = 1, \dots, \kappa\}.$$

(2) From Lemma 1, we know

$$\rho(\mathbf{N}) = \frac{1}{1 - \rho(\mathbf{T})},$$

then we get $\rho(\mathbf{N}_{\mathbf{q}}) \leq \max\{\rho(\mathbf{N}_{s_k}); k = 1, \dots, \kappa\}$. □ □

In the following we investigate whether and when the performance of a mixed strategy EA is better than a pure strategy EA.

Definition 3. A mutation operator s_1 is called complementary to another mutation operator s_2 on a fitness function $f(x)$ if for any x such that

$$P_{s_1}(x, x) = \rho(\mathbf{T}_{s_1}), \tag{10}$$

it holds

$$P_{s_2}(x, x) < \rho(\mathbf{T}_{s_1}). \tag{11}$$

Theorem 2. Let $f(x)$ be a fitness function and $EA(s_1)$ a pure strategy EA. If a mutation operator s_2 is complementary to s_1 , then it is possible to design a mixed strategy $EA(s_1, s_2)$ which satisfies

1. its asymptotic convergence rate is larger than that of $EA(s_1)$;
2. and its asymptotic hitting time is shorter than that of $EA(s_1)$.

Proof. (1) Design a mixed strategy EA(s_1, s_2) as follows. For any x such that

$$P_{s_1}(x, x) = \rho(\mathbf{T}_{s_1}),$$

let the strategy probability distribution satisfy

$$q_{s_2}(x) = 1.$$

For any other x , let the strategy probability distribution satisfy

$$q_{s_1}(x) = 1.$$

Because s_2 is complementary to s_1 , we get that

$$\rho(\mathbf{T}_{\mathbf{q}}) < \rho(\mathbf{T}_{s_1}),$$

and then

$$-\ln \rho(\mathbf{T}_{\mathbf{q}}) > -\ln \rho(\mathbf{T}_{s_1}),$$

which proves the first conclusion in the theorem.

(2) From Lemma 1

$$\rho(\mathbf{N}) = \frac{1}{1 - \rho(\mathbf{T})}$$

we get that

$$\rho(\mathbf{N}_{\mathbf{q}}) < \rho(\mathbf{N}_{s_k}), \quad \forall k = 1, \dots, \kappa,$$

which proves the second conclusion in the theorem. \square \square

Definition 4. κ mutation operators s_1, \dots, s_κ are called mutually complementary on a fitness function $f(x)$ if for any $x \in S_{\text{non}}$ and $sl \in \{s_1, \dots, s_\kappa\}$ such that

$$P_{sl}(x, x) \geq \min\{\rho(\mathbf{T}_{s_1}), \dots, \rho(\mathbf{T}_{s_\kappa})\}, \quad (12)$$

it holds: $\exists sk \neq sl$,

$$P_{sk}(x, x) < \min\{\rho(\mathbf{T}_{s_1}), \dots, \rho(\mathbf{T}_{s_\kappa})\}. \quad (13)$$

Theorem 3. Let $f(x)$ be a fitness function and s_1, \dots, s_κ be κ mutation operators. If these mutation operators are mutually complementary, then it is possible to design a mixed strategy EA which satisfies

1. its asymptotic convergence rate is larger than that of any pure strategy EA using one mutation operator;
2. and its asymptotic hitting time is shorter than that of any pure strategy EA using one mutation operator.

Proof. (1) We design a mixed strategy EA(s_1, \dots, s_κ) as follows. For any x and any strategy $sl \in \{s_1, \dots, s_\kappa\}$ such that

$$P_{sl}(x, x) \geq \min\{\rho(\mathbf{T}_{s_1}), \dots, \rho(\mathbf{T}_{s_\kappa})\},$$

from the mutually complementary condition, we know $\exists sk \neq sl$, it holds

$$P_{sk}(x, x) < \min\{\rho(\mathbf{T}_{s_1}), \dots, \rho(\mathbf{T}_{s_\kappa})\}.$$

Let the strategy probability distribution satisfy

$$q_{sk}(x) = 1.$$

For any other x , we assign a strategy probability distribution in any way.

Because the mutation operators are mutually complementary, we get that

$$\rho(\mathbf{T}_{\mathbf{q}}) < \min\{\rho(\mathbf{T}_{s_1}), \dots, \rho(\mathbf{T}_{s_\kappa})\},$$

and then

$$-\ln \rho(\mathbf{T}_{\mathbf{q}}) > \min\{-\ln \rho(\mathbf{T}_{s1}), \dots, -\ln \rho(\mathbf{T}_{s\kappa})\},$$

which proves the first conclusion in the theorem.

(2) From Lemma 1

$$\rho(\mathbf{N}) = \frac{1}{1 - \rho(\mathbf{T})},$$

we get that

$$\rho(\mathbf{N}_{\mathbf{q}}) < \rho(\mathbf{N}_{sk}), \quad \forall k = 1, \dots, \kappa,$$

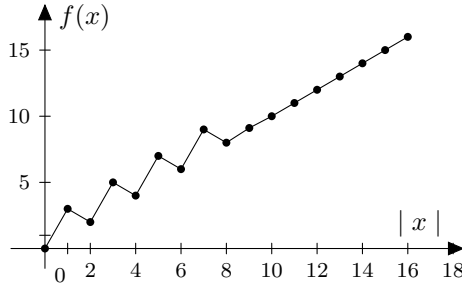
which proves the second conclusion in the theorem. \square \square

Example 3. Consider the problem of maximizing the following fitness function $f(x)$ (see Figure 2):

$$f(x) = \begin{cases} |x|, & \text{if } |x| < 0.5n \text{ and } |x| \text{ is even;} \\ |x| + 2, & \text{if } |x| < 0.5n \text{ and } |x| \text{ is odd;} \\ |x|, & \text{if } |x| \geq 0.5n. \end{cases}$$

where $x = (b_1 \dots b_n)$ is a binary string, n the string length and $|x| := \sum_{i=1}^n b_i$.

Figure 2: The shape of the function $f(x)$ in Example 3 when $n = 16$.



Consider two common mutation operators:

- $s1$: to choose one bit randomly and then flip it;
- $s2$: to flip each bit independently with a probability $1/n$.

$EA(s1)$ uses the mutation operator $s1$ only. Then $\rho(\mathbf{T}_{s1}) = 1$, and then the asymptotic convergence rate is $R(\mathbf{T}_{s1}) = 0$.

$EA(s2)$ utilizes the mutation operator $s2$ only. Then

$$\rho(\mathbf{T}_{s2}) = 1 - \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1}.$$

We have

$$\min\{\rho(\mathbf{T}_{s1}), \rho(\mathbf{T}_{s2})\} = 1 - \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1}.$$

(1) For any x such that

$$P_{s1}(x, x) \geq 1 - \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1},$$

we have

$$P_{s1}(x, x) = 1,$$

and we know that

$$P_{s2}(x, x) < 1 - \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1}.$$

(2) For any x such that

$$P_{s2}(x, x) = \rho(\mathbf{T}_{s2}) = 1 - \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1},$$

we know that

$$P_{s1}(x, x) = 1 - \frac{1}{n} < \rho(\mathbf{T}_{s2}) = 1 - \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1}.$$

Hence these two mutation operators are mutually complementary.

We design a mixed strategy $EA(s1, s2)$ as follows: let the strategy probability distribution satisfy

$$q_{s1}(x) = \begin{cases} 0, & \text{if } |x| \leq 0.5n; \\ 1, & \text{if } |x| > 0.5n. \end{cases}$$

According to Theorem 3, the asymptotic convergence rate of this mixed strategy $EA(s1, s2)$ is larger than that of either $EA(s1)$ or $EA(s2)$.

5 Conclusion and Discussion

The result of this paper is summarized in three points.

- Asymptotic convergence rate and asymptotic hitting time are proposed to measure the performance of EAs. They are seldom used in evaluating the performance of EAs before.
- It is proven that the asymptotic convergence rate and asymptotic hitting time of any mixed strategy (1+1) EA consisting of several mutation operators is not worse than that of the worst pure strategy (1+1) EA using only one of these mutation operators.
- Furthermore, if these mutation operators are mutually complementary, then it is possible to design a mixed strategy EA whose performance (asymptotic convergence rate and asymptotic hitting time) is better than that of any pure strategy EA using one mutation operator.

An argument is that several mutation operators can be applied simultaneously, e.g., in a population-based EA, different individuals adopt different mutation operators. However in this case, the number of fitness evaluations at each generation is larger than that of a (1+1) EA. Therefore a fair comparison should be a population-based mixed strategy EA against a population-based pure strategy EA. Due to the length restriction, this issue will not be discussed in the paper.

Acknowledgement:

J. He is partially supported by the EPSRC under Grant EP/I009809/1. H. Dong is partially supported by the National Natural Science Foundation of China under Grant No. 60973075 and Natural Science Foundation of Heilongjiang Province of China under Grant No. F200937, China.

References

- [1] D.B. Fogel and Z. Michalewicz. *Handbook of Evolutionary Computation*. Oxford Univ. Press, 1997.
- [2] C. Grosan, A. Abraham, and H. Ishibuchi. *Hybrid Evolutionary Algorithms*. Springer Verlag, 2007.
- [3] P.K. Dutta. *Strategies and Games: Theory and Practice*. MIT Press, 1999.
- [4] J. He and X. Yao. A game-theoretic approach for designing mixed mutation strategies. In L. Wang, K. Chen, and Y.-S. Ong, editors, *Proceedings of the 1st International Conference on Natural Computation, LNCS 3612*, pages 279–288, Changsha, China, August 2005. Springer.

- [5] H. Dong, J. He, H. Huang, and W. Hou. Evolutionary programming using a mixed mutation strategy. *Information Sciences*, 177(1):312–327, 2007.
- [6] L. Shen and J. He. A mixed strategy for evolutionary programming based on local fitness landscape. In *Proceedings of 2010 IEEE Congress on Evolutionary Computation*, pages 350–357, Barcelona, Spain, July 2010. IEEE Press.
- [7] R.S. Varga. *Matrix Iterative Analysis*. Springer, 2009.
- [8] J. He and T. Chen. Population scalability analysis of abstract population-based random search: Spectral radius. *Arxiv preprint arXiv:1108.4531*, 2011.
- [9] Z. Michalewicz. *Genetic Algorithms + Data Structure = Evolution Program*. Springer Verlag, New York, 1996.
- [10] J. He and Y. Zhou. A comparison of GAs using penalizing infeasible solutions and repairing infeasible solutions II: Avarerage capacity knapsack. In L. Kang, Y. Liu, and S. Y. Zeng, editors, *Proceedings of the 2nd International Symposium on Intelligence Computation and Applications, LNCS 4683*, pages 102–110, Wuhan, China, September 2007. Springer.
- [11] G. Rudolph. Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 5(1):96–101, 1994.
- [12] J. He and L. Kang. On the convergence rate of genetic algorithms. *Theoretical Computer Science*, 229(1-2):23–39, 1999.
- [13] J. He and X. Yao. Towards an analytic framework for analysing the computation time of evolutionary algorithms. *Artificial Intelligence*, 145(1-2):59–97, 2003.
- [14] M. Iosifescu. *Finite Markov Chain and their Applications*. Wiley, Chichester, 1980.